

# ICA APLICADO A LA EXTRACCIÓN DE CARACTERÍSTICAS EN IMÁGENES

## ICA applied to image feature extraction

### RESUMEN

La extracción de características es una de las tareas claves que deben resolver los sistemas de visión artificial que cada vez presentan nuevos campos de aplicación. En este documento se exponen los fundamentos matemáticos del análisis de componentes independientes (ICA), se muestra su aplicación como técnica de caracterización de imágenes y se valida en la tarea de reconocimiento de rostros sobre las bases de datos CMU y ORL.

**PALABRAS CLAVES:** Extracción de características, análisis de componentes independientes, reconocimiento de rostros, visión artificial.

### ABSTRACT

*The feature extraction is one of the key tasks to be solved by machine vision systems that increasingly present new fields of application. These documents sets out the mathematical foundation of the independent component analysis (ICA), show its implementation such as image characterization technique and is validated in the task of face recognition on the CMU and ORL databases.*

**KEYWORDS:** *feature extraction, independent component analysis, face recognition, machine vision.*

## 1. INTRODUCCIÓN

Los sistemas de visión artificial actuales necesitan extraer información de las imágenes y requieren que el procesamiento de estas se realice con una alta velocidad de forma que se pueda analizar gran cantidad de información.

Actualmente esta tarea no ha sido resuelta satisfactoriamente, debido a que se utilizan técnicas que encuentran representaciones no apropiadas que no permiten identificar acertadamente patrones de interés de las imágenes porque extraen conjuntos con muchas características que pueden estar correlacionadas.

Uno de los métodos de extracción de características más usados es el análisis de componentes principales (PCA), que se basa en características estadísticas para encontrar una base representacional donde las características (estadísticas de segundo orden) no están correlacionadas. Pero tiene como desventaja que puede no ser apropiada para el reconocimiento de determinados patrones (por ejemplo rostros y expresiones faciales) en donde la información de importancia está contenida en las estadísticas de alto orden [1].

Recientemente el ICA, el cual fue concebido para el problema de separación ciega de fuentes (proceso que se fundamenta en la independencia estadística de las señales, que son generadas en procesos independientes y consiste en la recuperación de las señales originales a

### DAMIAN A. ALVAREZ

Ingeniero Electricista.  
Profesor Catedrático  
Universidad Tecnológica de Pereira  
[damianalvarez@ohm.utp.edu.co](mailto:damianalvarez@ohm.utp.edu.co)

### EDUARDO GIRALDO

M.Sc. Ingeniero Electricista.  
Profesor Auxiliar  
Universidad Tecnológica de Pereira  
[egiraldos@utp.edu.co](mailto:egiraldos@utp.edu.co)

partir de las mezclas observadas), ha recibido gran atención en redes neuronales y procesamiento de señales, debido al potencial de sus aplicaciones en estas áreas [1]. De igual forma se ha extendido su aplicación al área de procesamiento de imágenes contemplando el hecho de que esta técnica puede ser aplicada en cualquier dominio donde un arreglo de  $N$  entradas forma una combinación lineal de un conjunto desconocido de fuentes estadísticamente independientes. Estudios como [1], [2] y [3] consideran de utilidad aplicar ICA en imágenes.

Según [4] aplicaciones reales en imágenes con frecuencia prefieren ICA sobre PCA debido a que ICA es capaz de producir bases no ortonormales y no está restringido a mediciones de covarianza como PCA. Sin embargo plantean usar PCA para estimar una matriz de mezclas que pueda ser usada como inicialización para ICA. Adicionalmente con ICA se encuentran una representación robusta de la imagen en la cual tanto las estadísticas de segundo como las de más alto orden están decorrelacionadas [2].

## 2. ANTECEDENTES

Existen dos diferentes campos de investigación que utilizan ICA para resolver el problema de separación ciega de fuentes. Por una parte están los estudios propuestos en 1991 por Jutten-Herault [5], continuados en 1994 y 1996 por Comon [6] y [7] y Cichocki [8] y [9].

De otra forma están los estudios basados en reglas de aprendizaje no supervisado, propuestas en 1992 por

Linsker, continuados en 1994 por Laheld [10] y en 1996 por Cardoso [11] y Amari [12] y explicados por primera vez desde el punto de vista de teoría de la información en 1995 por Bell y Sejnowski [13] quienes aplicaron con éxito su algoritmo para realizar la separación de fuentes independientes mezcladas aleatoriamente y para revertir el efecto de un filtro desconocido.

La regla de aprendizaje de [13] conocida como *Infomax* es aplicable tan solo para fuentes Super- Gaussianas. Sin embargo Girolami y Fyfe [14] derivaron una regla de aprendizaje que hace posible la separación ciega de mezclas de fuentes con distribuciones sub y super Gaussianas. En [15] se muestra que la regla de aprendizaje es una extensión de [13] satisfaciendo un criterio de estabilización y que puede optimizarse para una mejor convergencia usando el gradiente natural o el relativo. Simulaciones en señales fisiológicas muestran la potencia de los métodos propuestos [15].

En los modelos ICA los vectores aleatorios estadísticamente independientes tienen distribuciones no Gaussianas y están mutuamente decorrelacionadas. Pero en [13] ni la matriz de mezclas ni su pseudoinversa son ortogonales, esta restricción para que la transformación sea decorrelacionada (los vectores independientes sean una transformación ortogonal de los datos) es realizada en forma aproximada, sin embargo en 1997 Hyvärinen en [16] y [17] introdujo el algoritmo *FastICA* el cual realiza una restricción exacta, usando un esquema de iteración denominado punto fijo, para optimizar la kurtosis. Con este algoritmo se consiguió extender el campo de análisis a problemas de mayor dimensionalidad. Actualmente las investigaciones en el área siguen realizándose, para mayor información se puede remitir a [8], [11], [12].

### 3. ICA

#### 3.1 Motivación

Si se imagina a  $n$  personas en un salón hablando simultáneamente y se tienen diferentes micrófonos ubicados en distintas locaciones. Los micrófonos darán grabaciones de tiempo  $x_1(t)$ ,  $x_2(t)$ ,  $x_n(t)$ , las cuales serán sumas ponderadas de las señales de voz, denotadas por  $s_1, s_2, \dots, s_n$ . Como se expresan en la ecuación lineal (1):

$$\begin{aligned} x_1(t) &= a_{11}s_1 + a_{12}s_2 + \dots + a_{1n}s_n \\ x_2(t) &= a_{21}s_1 + a_{22}s_2 + \dots + a_{2n}s_n \\ &\vdots \\ x_n(t) &= a_{n1}s_1 + a_{n2}s_2 + \dots + a_{nn}s_n \end{aligned} \quad (1)$$

donde los  $a$  son parámetros que dependen de la distancia de los micrófonos a las personas. Este problema que consiste en estimar las señales de voz originales  $s_1, s_2, \dots, s_n$  a partir de las grabaciones  $x_1(t)$ ,  $x_2(t)$ ,  $x_n(t)$ , es conocido como el problema de la fiesta de cocteles.

Para resolver problemas estrechamente relacionadas con la separación ciega de fuentes, se ha desarrollado ICA, técnica que ha tenido un creciente interés y a la cual se le ha encontrado gran cantidad de aplicaciones (como lo es por ejemplo la extracción de características).

#### 3.2 Definición

El problema que intenta solucionar ICA puede formularse en notación matricial, mediante la ecuación:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \quad (2)$$

donde  $A$  es la denominada matriz de mezcla (desconocida para nosotros).

Consiste en recuperar el vector de fuentes  $S$ , disponiendo, únicamente, de un vector de  $N$  observaciones  $X$ ; es decir, el objetivo es llegar a calcular una matriz de pesos  $W$ , la cual es la inversa de  $A$  con algunos cambios en escala y permutaciones, de modo que se pueda recuperar las señales independientes originales como:

$$\mathbf{U} = \mathbf{W}\mathbf{X} \quad (3)$$

#### 3.3 Hipótesis

Para el estudio de la separación ciega de fuentes a partir de mezclas se suelen establecer las siguientes hipótesis:

1. Las fuentes son estadísticamente independientes.
2. Las fuentes tienen distribución no-Gaussiana, es decir como máximo sólo una de ellas puede ser Gaussiana.
3. El modelo de mezcla se conoce.

Para conocer si se cumple la primera hipótesis es necesario definir que es independencia

#### 3.4 Independencia

Para esto se consideran dos valores escalares aleatorios de las variables  $y_1$  y  $y_2$ . Básicamente las variables  $y_1$  y  $y_2$  son independientes si la información sobre el valor de  $y_1$  no contiene información acerca del valor de  $y_2$  y viceversa. Esto se define en términos de la función de densidad de probabilidad. Denotando  $p(y_1, y_2)$  como la función de densidad de probabilidad conjunta y  $p_1(y_1)$  y  $p_2(y_2)$  como funciones de densidad de probabilidad marginales [18]:

$$p_1(y_1) = \int p(y_1, y_2) dy_2 \quad p_2(y_2) = \int p(y_1, y_2) dy_1 \quad (4)$$

Se dice que  $y_1$  y  $y_2$  son independientes si y solo si la densidad de probabilidad conjunta es factorizable en la siguiente forma [18]:

$$p_1(y_1, y_2) = p_1(y_1)p_2(y_2) \quad (5)$$

Definición que puede extenderse naturalmente para cualquier número de variables aleatorias.

Para la verificación de independencia de las fuentes se suele recurrir a la minimización de la información mutua o los momentos y los cumulantes [18].

### 3.5 Medidas de no Gaussianidad

Intuitivamente hablando la clave para estimar un modelo ICA es la no Gaussianidad. Actualmente sin no Gaussianidad la estimación no es posible.

Para usar la no Gaussianidad en la estimación de ICA se debe tener una medida cuantitativa de no Gaussianidad de una variable aleatoria dada  $y$ , que se asume es centrada (tiene media cero) y su varianza es igual a uno. Lo que se consigue incluyendo una etapa de preprocesamiento al algoritmo ICA como se verá posteriormente.

#### 3.5.1 Kurtosis

La kurtosis es una estadística de cuarto orden. Se define para una variable aleatoria  $y$  de media cero como [18]:

$$kurt(y) = E\{y^4\} - 3[E\{y^2\}]^2 \quad (6)$$

Si se asume que  $y$  tiene varianza unitaria, sería entonces:

$$kurt(y) = E\{y^4\} - 3 \quad (7)$$

La kurtosis sin embargo tiene como problema que es sensible a factores externos. Su valor puede depender de unas pocas observaciones de la distribución que pueden ser irrelevantes. Es decir no representa una medición robusta de la no Gaussianidad [18].

#### 3.5.2 Negentropía

Una segunda e importante medición de no Gaussianidad es dada por la negentropía que se calcula basándose en la teoría de información a partir de la entropía, la cual puede ser interpretada como el grado de información que la observación da de la variable (entre más aleatoria sea una variable mayor será su entropía). La entropía diferencial  $H$  para una variable aleatoria  $y$  de valores continuos, se define como [18]:

$$H(y) = \int f(y) \log f(y) dy \quad (8)$$

Con el fin de obtener una medición de no Gaussianidad que sea cero para una variable Gaussiana y sea siempre

no negativa, se usa una versión ligeramente modificada conocida como negentropía, la cual se define como [18]:

$$J(y) = H(y_{gauss}) - H(y) \quad (9)$$

donde  $y_{gauss}$  es un vector Gaussiano aleatorio de la misma matriz de covarianza de  $y$ .

La ventaja de usar la negentropía es que está mejor justificada estadísticamente, su problema sin embargo es que es más difícil de estimar, por lo que generalmente se usan aproximaciones, que pueden utilizar momentos de alto orden, así [18]:

$$J(y) \approx \frac{1}{12} E\{y^3\}^2 + \frac{1}{48} kurt\{y\}^2 \quad (10)$$

Debido a que esta aproximación resulta algo limitada, es necesario recurrir a otras formas de aproximación tal como el principio de máxima entropía, planteado por Hyvärinen siguiendo la expresión (14) [18].

## 4. ALGORITMOS ICA

### 4.1 INFOMAX

Una de las primeras formas planteadas para estimar ICA fue el algoritmo de aprendizaje no supervisado **Infomax**, propuesto por Bell y Sejnowski [13], el cual se basa en la principio de transferencia de información óptima a través de neuronas.

El algoritmo no necesita ningún conocimiento previo de la distribución de las entradas, es decir es una regla de aprendizaje no supervisada. Está definido para señales que no estén corruptas por ningún tipo de ruido y para trabajar con fuentes Super-Gaussianas

En este algoritmo se maximiza la información mutua que la salida de una red neuronal contiene acerca de la entrada para maximizar indirectamente la información mutua entre unidades entrada; esto se consigue cuando se maximiza la entropía de la salida dada por:

$$H(y) = -E[\ln(f_y(y))] = -\int_{-\infty}^{\infty} f_y(y) \ln(f_y(y)) dy \quad (11)$$

donde  $f_y(y)$  es la función de densidad de probabilidad de la salida, la cual para el caso de múltiples entradas y salidas esta dada por:

$$f_y(y) = \frac{f_x(x)}{|\mathbf{J}|} \quad (12)$$

donde  $\mathbf{J}$  es el Jacobiano de la transformación y esta dado por el determinante de la matriz de derivadas parciales de la salida  $\mathbf{y}$ .

En el algoritmo *Infomax* la maximización de la entropía dada por (11) se realiza usando gradiente ascendente sobre los elementos de  $\mathbf{W}$ . Para obtener una regla de actualización de pesos dada por:

$$\Delta \mathbf{W} = (\mathbf{I} + \mathbf{Y}^T \mathbf{U}^T) \mathbf{W} \quad (13)$$

donde  $\mathbf{I}$  es la matriz identidad.

#### 4.2 FastICA

En [16], Hyvärinen, introdujo un algoritmo ICA que cumple con las restricciones de ortogonalidad en forma exacta que de acuerdo a [19] permite realizar una transformación ortonormal de los datos; como este algoritmo presentó una convergencia más rápida se le denominó *FastICA*.

El algoritmo *FastICA* usa una nueva función objetivo para estimar la transformación ICA, a diferencia de la aproximación realizada en [13], este algoritmo es más general, ya que maximiza cualquier función de costo de la forma de la expresión (14) y no solo la entropía, para maximizar la independencia.

$$J = c[E\{G(y_i)\} - E\{G(v)\}^2] \quad (14)$$

donde  $G$  es una función no cuadrática que en la práctica es elegida de forma que sea rápida de calcular y que considere el orden en que las componentes son estimadas,  $c$  es una constante arbitraria,  $v$  es una variable Gaussiana aleatoria,  $y_i$  también es una variable aleatoria y ambas se asumen de media cero y varianza unitaria.

Lo primero que se debe hacer es encontrar las componentes independientes maximizando la expresión:

$$J_G(\mathbf{W}) = \left[ E\{E(\mathbf{W}^T \mathbf{X})\} - E\{G(v)\} \right]^2 \quad (15)$$

Donde  $\mathbf{W}$  es un vector de pesos que debe cumplir con:

$$E\left\{\left(\mathbf{W}^T \mathbf{X}\right)^2\right\} = 1 \quad (16)$$

Lo segundo, es extender el uso de una función objetivo unitaria para calcular todo el arreglo  $\mathbf{W}$ . Para hacer esto la información mutua es minimizada cuando la suma de la negentropía de las componentes transformadas es maximizada, además se considera la restricción de no correlación para obtener el problema:

$$\max \sum_{i=1}^n J_G(\mathbf{w}_i) \text{ con } E\left\{\left(\mathbf{w}_k^T \mathbf{X}\right)\left(\mathbf{w}_j^T \mathbf{X}\right)\right\} = \delta_{kj} \quad (17)$$

Maximizando (17) se obtiene cada uno de los vectores  $\mathbf{w}_j$  que conforma cada una de las filas de la matriz  $\mathbf{W}$ , con la cual se calcula la transformación ICA. Esta maximización es llevada a cabo usando un simple y eficiente esquema de iteración denominado algoritmo de punto fijo que tiene propiedades de convergencia deseables y encuentra componentes independientes de cualquier distribución no-Gaussiana [16].

### 5. PREPROCESAMIENTO PARA ICA

Antes de aplicar algún algoritmo ICA, es útil hacer un preprocesamiento, para hacer el problema de estimación de ICA más simple y mejor condicionado [18].

#### 5.1 Centrado

El preprocesamiento más básico es centrar  $\mathbf{X}$ , esto se logra por ejemplo sustrayendo su vector de media para hacer que  $\mathbf{X}$  sea una variable de media cero. Este paso es realizado con el fin simplificar los algoritmos ICA, no quiere decir que se no puede dejar de hacer [18].

#### 5.2 Blanqueamiento

Otra estrategia de preprocesamiento útil es primero blanquear las variables observadas. Esto significa que antes de aplicar algún algoritmo ICA y después del centrado, se transforma el vector observado  $\tilde{\mathbf{X}}$ , para obtener un nuevo vector blanqueado  $\tilde{\mathbf{X}}$ , con componentes no correlacionadas y sus varianzas iguales a la unidad. Un método común de blanqueamiento es sustraer las filas de  $\tilde{\mathbf{X}}$  y seguidamente pasar  $\tilde{\mathbf{X}}$  a través de una matriz de blanqueamiento dada por [18]:

$$\mathbf{W}_z = 2 \left( \text{Cov}(\mathbf{X}) \right)^{-\frac{1}{2}} = 2 \left( \mathbf{X} \mathbf{X}^T \right)^{-\frac{1}{2}} \quad (18)$$

Esta etapa de blanqueamiento reduce las estadísticas de primer y segundo orden de los datos y tiene como principal ventaja que entrega una nueva matriz de mezclas ortogonal [18].

Cuando las entradas a ICA son los datos que han pasado por este paso previo, la matriz de transformación es  $\mathbf{W}_i = \mathbf{W} \mathbf{W}_z$ .

### 6. APLICACIÓN DE ICA A IMÁGENES

La utilización de ICA en imágenes fue propuesta por Bartlett [2] en donde se plantea que esta aplicación puede hacerse por medio de dos diferentes arquitecturas: tratando las imágenes como variables aleatorias y los

píxeles como observaciones o salidas (ver figura 1) ó tratando los píxeles como variables aleatorias y las imágenes como salidas. La primera arquitectura produce imágenes bases estadísticamente independientes, mientras la segunda arquitectura usa ICA para encontrar una representación en la cual los coeficientes usados para codificar las imágenes son estadísticamente independientes, básicamente es una modificación de la arquitectura I, la cual consiste en transponer la matriz de entrada  $X$  de forma que las imágenes sean columnas y los píxeles filas, como se muestra en la figura 2 [2].

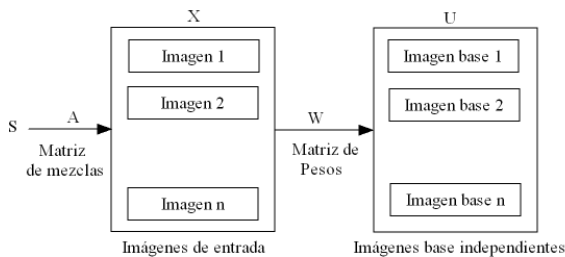


Fig. 1. Arquitectura I para aplicar ICA en imágenes.

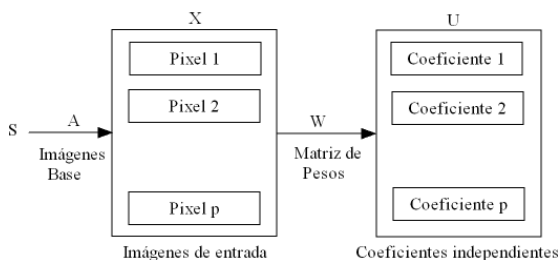


Fig. 2. Arquitectura II para aplicar ICA en imágenes.

En la arquitectura II, cada imagen de una base de datos es considerada una combinación lineal de las imágenes bases dadas por la matriz  $A$ . Cada imagen base es asociada con un conjunto de causas, dado por un vector de coeficientes en  $S$ . Las causas son recobradas por una matriz de pesos  $W$  (encontrados con ICA), los cuales producen salidas  $U$  estadísticamente independientes.

En [2] se plantea un método para encontrar un conjunto de filtros que producen salidas  $U$  mediante la expresión (3), a partir de los pesos  $W$  obtenidos con la expresión (13), se usa PCA como método de inicialización, se incluye una etapa de preprocesamiento, tal como en [18] y se utiliza  $U$  como entrada al clasificador del vecino más cercano. En este documento se implementará este método con el fin de hacer reconocimiento de rostros. Para la validación de los resultados se usará el esquema de validación cruzada seleccionando conjuntos del 70% y 30% para el entrenamiento y prueba respectivamente.

Se desea probar lo expuesto en [20], en donde se plantea que la arquitectura II es más apropiada para tareas holísticas (p.ej reconocimiento de rostros), debido a que entrega vectores de características globales. Mientras que la arquitectura I es mejor para tareas localizadas (p.ej reconocimiento de expresiones), ya que produce

características localizadas que solamente son influenciadas por pequeñas partes de la imagen.

## 7. RESULTADOS Y DISCUSIÓN

Se hizo indispensable realizar inicialmente un acondicionamiento de las imágenes a través de ecualización del histograma con el fin de homogenizar las condiciones de iluminación. La mejor forma de aplicar ICA a las imágenes para reconocer rostros fue mediante la arquitectura II. Se identificó a la totalidad de los sujetos de las bases de datos CMU [21] y ORL [22] (que tienen los rostros previamente detectados), usando todas las imágenes de dichas bases de datos. Los mejores resultados se obtuvieron utilizando el algoritmo *Infomax*, alcanzando un acierto de  $100.00\% \pm 0.11$  para la base de datos CMU y  $88.58\% \pm 0.24$  para la ORL.

En este documento se mostraron los fundamentos de los algoritmos *Infomax* y *FastICA* que han sido usados para la separación ciega de fuentes y la extracción de características. De acuerdo a estos fundamentos y los expuestos en trabajos como [19] los algoritmos vistos maximizan funciones con el mismo óptimo global y por lo tanto sus resultados deben converger a la misma solución, pero como en la práctica las diferentes formulaciones utilizan diferentes técnicas de aproximación, se tiene que los algoritmos encuentran distintas soluciones. En el experimento realizado se verificó que existen diferencias en el desempeño entre estos algoritmos, sin embargo estas dependieron más de las bases de datos a utilizar, siendo más pequeñas las diferencias para la base de datos ORL que para la CMU.

En general la base de datos ORL presentó mayor dificultad para ICA debido a que posee variaciones de pose en donde existe en mayor proporción relaciones entre píxeles de tipo no lineal, comparado con la base de datos CMU que tiene cambios de expresiones faciales en donde existe menores relaciones de tipo no lineal.

## 8. CONCLUSIONES

La transformación ICA es una técnica apropiada para la extracción de característica en imágenes ya que ofrece una representación robusta en donde las características están lo más decorrelacionadas posible. Permitió el reconocimiento de rostros con porcentajes de acierto promedio del  $100\% \pm 0.11$  y  $88.56\% \pm 0.24$  para las base de datos CMU y ORL respectivamente.

Se comprobó que el desempeño de ICA está relacionado al tipo de arquitectura sobre la cual se aplica a las imágenes ya que estas entregan distintos tipos de información, considerando esto se debe ser cuidadoso en la selección de la arquitectura, ya que esta cambia dependiendo de la naturaleza de la tarea a resolver.

## 9. ENFOQUES FUTUROS

Aunque la transformación ICA encuentra direcciones independientes o un conjunto de características lo mayor decorrelacionado posible, es necesario hacer una selección de las características a usar en la construcción del clasificador de forma que no permanezcan características aun correlacionadas que no aportan información. Como trabajo futuro se plantea utilizar ICA como técnica primaria de caracterización acompañada de métodos de selección de características que permitan que aquellas componentes con más alto grado de separación emerjan más fácilmente que las otras, lo cual incrementaría la probabilidad de que las componentes seleccionadas sean más significantes en la clasificación.

Como en los distintos algoritmos ICA, los vectores independientes garantizan una transformación ortogonal de los datos [19], es decir la transformación ICA realiza una rotación de los datos originales es conveniente utilizar clasificadores invariantes a rotación (p.ej Naive-Bayes) en donde existen ventajas al utilizar la rotación que hace ICA. Mientras si se utilizan clasificadores no invariantes a la rotación (aquellos basados en distancias), no se aprovecharían las ventajas de ICA y bajo estas circunstancias puede no superarse los resultados de PCA.

Los algoritmos de ICA expuestos en este documento son intrínsecamente lineales, por lo que para la extracción de características pueden no representar en forma adecuada los datos con estructuras no lineales. Una forma de sobrellevar este problema sería realizar primero un mapeo no lineal de los datos en un espacio característico utilizando métodos Kernel y aplicar ICA sobre este espacio transformado.

## 10. AGRADECIMIENTOS

Esta investigación fue realizada gracias al apoyo del grupo de Control e Instrumentación de la UTP en el marco del proyecto COLCIENCIAS 111037019600.

## 11. BIBLIOGRAFÍA

[1] N. Kwak, "Feature extraction for classification problems and its application to face recognition", *Pattern Recognition*, vol. 41, no. 5, pp. 1718-1734, 2008.  
 [2] M. S. Bartlett, "Face image analysis by unsupervised learning and redundancy reduction," Ph.D. dissertation, 1998, chair-Donald I. Macleod.  
 [3] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying facial actions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 10, pp. 974-989, 1999.  
 [4] Yun Q. Shi and Huiyang Sun. "Multimedia image and video processing", CRC Press LLC, 2001.  
 [5] C. Jutten and J. Herault, "Blind separation of sources. An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, vol. 24, pp. 1-10, 1991.

[6] P. Comon, C. Jutten, and J. Herault, "Blind separation of sources. Problems statement," *Signal Processing*, vol. 24, no. 1, pp. 11-20, 1991.  
 [7] P. Comon "Independent Component Analysis, a new concept " *Signal Processing*, Elsevier, vol. 36, no. 3, pp. 287-314, 1994, special issue on Higher-Order Statistics.  
 [8] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust neural networks with on-line learning for blind identification and blind separation of sources," *Electronics Letters*, vol. 30, no. 17, pp. 1386-1387, 1994.  
 [9] A. Cichocki and W. Kasprzak, "Blind separation of sources. 2. problems statement," *Neural network world*, vol. 6, no. 4, pp. 515-523, 1996.  
 [10] B. Laheld and J. Cardoso, "Adaptive source separation with uniform performance," 1994.  
 [11] J.-F. Cardoso, "Equivariant adaptive source separation," *IEEE Trans. on S.P.*, vol. 44, no. 45, pp. 3017-3030, 1996.  
 [12] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, vol. 8. The MIT Press, 1996, pp. 757-763.  
 [13] A.J. Bell and T.J. Sejnowski. "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, 7(6), p.p. 1129-1159, 1995.  
 [14] M. Girolami and C. Fyfe, "Negentropy and kurtosis as projection pursuit indices provide generalised ica algorithms," 1997.  
 [15] T.-W. Lee, M. Girolami, A. J. Bell, and T. J. Sejnowski, "A unifying information-theoretic framework for independent component analysis," *Int. journal of computers and mathematics with applications*, 1999.  
 [16] A. Hyvärinen, "The fixed-point algorithm and maximum likelihood estimation for independent component analysis," *Neural Processing Letters*.  
 [17] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Transactions on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.  
 [18] Salua Esther Nassabay. "Estudio comparativo de la técnica Análisis de Componentes Independientes ICA aplicado al procesamiento digital de imágenes con ruido", tesis doctoral, 2006, Universidad de Granada.  
 [19] M. A. Vicente, P. O. Hoyer, and A. Hyvärinen, "Equivalence of some common linear feature extraction techniques for appearance-based object recognition tasks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 5, pp. 896-900, 2007.  
 [20] B. A. Draper, K. Baek, M. S. Bartlett, and J. R. Beveridge, "Recognizing faces with pca and ica," *Comput. Vis. Image Underst.*, vol. 91, no. 1-2, pp. 115-137, 2003.  
 [21] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression database," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615-1618, 2003.  
 [22] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," 1994.